ランダム行列との比較によるNYSE株価1時間変動の相関行列解析(経済物理学とその周辺,統計数理研究所研究会共同研究集会,経済物理学2009-ミクロとマクロの架け橋-,京都大学基礎物理学研究所2009年度前期研究会,研究会報告)

AUTHOR(S):

田中, 美栄子; 田中, 瑤子; 伊藤, 大哲; 中村, 元紀; 木戸, 丈剛; 河村, 綾; 佐藤, 彰洋

CITATION:

田中, 美栄子 ...[et al]. ランダム行列との比較によるNYSE株価1時間変動の相関行列解析(経済物理学とその周辺,統計数理研究所研究会共同研究集会,経済物理学2009-ミクロとマクロの架け橋-,京都大学基礎物理学研究所2009年度前期研究会,研究会報告). 物性研究 2010, 93(5): 675-676

ISSUE DATE:

2010-02-05

URL:

http://hdl.handle.net/2433/169224

ランダム行列との比較によるNYSE株価1時間変動の相関行列解析

田中 美栄子†, 田中 瑶子, 伊藤大哲, 中村元紀, 木戸丈剛, 河村綾, 佐藤彰洋*
鳥取大学大学院工学研究科†,鳥取大学工学部知能情報工学科,京都大学大学院情報学研究科*

1. はじめに

株価のようにランダム性の強い時系列から有意成分を分離する方法として,時系列間の同時刻相関行列のスペクトルを,ランダム行列から作った相関行列のスペクトルからの外れ成分として扱う方法があり,様々の分野に応用されているが,金融時系列に対してはNYSE株価のうちS&P500に入っているものを選んでその日次変動を扱ったPlerou等の研究[1]が知られている.本研究はNYSE株価の内,取引量の多い419社の株価1時間変動を用いて同様の解析を行った結果を報告する.

2. 株価tickデータの同時刻相関行列

等間隔でないtickデータから同時刻相関行列を作るため,定時に最も近い時点での価格を代表値として用いる.市場開始の9時半から半時間後の10時から始めて,1時間毎に1日あたり6点が採れる.1994年の営業日 252 日分で長さ T=1512の時系列を N=419 社に対して抽出した時系列の行列 (i=1,..N; k=1,..,T):

$$x_{i,k} = \begin{vmatrix} x_{1,1} & x_{1,2} & \cdot & x_{i,T} \\ x_{2,1} & x_{2,2} & \cdot & x_{2,T} \\ \cdot & \cdot & \cdot & \cdot \\ x_{N,1} & x_{N,2} & \cdot & x_{N,T} \end{vmatrix} \qquad (1)$$

から作った相関行列

$$C_{i,j} = \frac{1}{T}\sum_{k=1}^{T} x_{i,k}x_{j,k} \qquad (2)$$

の固有値スペクトルを,ランダム行列理論のよく知られた結果と比較して有意成分を分離する.

2-1. 変数の確認

株価 $S_{i,k}$ から作った対数収益率

$$X_{i,k} = \ln(S_{i,k+1}/S_{i,k}) \approx \Delta S_{i,k}/S_{i.k} \qquad (3)$$

から平均値を差し引き標準偏差で割って正規化した変数

$$x_{i,k} = \frac{X_{i,k} - \langle X_i \rangle}{\sigma} \qquad (4)$$

を(1)-(2)式の変数として用いる.この変数の頻度分布は正規分布と非常に近く,ランダム行列理論と比較するのに良い変数であると言える.Fig.1に1994年のNYSE-TAQ(trade)の株価時系列から作成した正規化対数収益率（角破線）とその最適関数(破線)と標準正規分布（実線）の比較を示す.

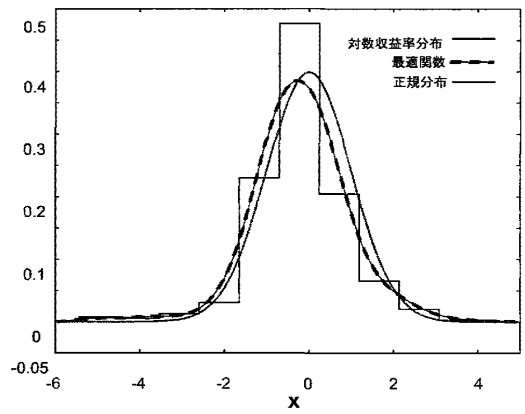

Fig.1 NYSE株価419銘柄から採った(4)式の変数の頻度分布,その最適関数,正規分布の比較

2-2. 相関行列の固有値スペクトル

上記の手続きで(2)式により計算した相関行列 C の固有値λと固有ベクトル U は

$$CU = \lambda U \qquad (5)$$

の関係を満たす.CはN×Nの成分を持ち,N=419と大きいので数値計算で対角化する.Cは対称行列であるから Jacobi 回転を使って初等的な方法で対角化できる.

ランダム行列の一般的性質はよく研究されており,ランダム時系列から作った相関行列の固有値スペクトル分布は

$$N \to \infty,\ T \to \infty,\ Q = T/N = \text{const.} \tag{6}$$

の極限で

$$P_{rm}(\lambda) = \frac{Q}{2\pi}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \tag{7}$$

となる. ここで分布の両端点は以下のようになる.

$$\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}},\ \ \lambda_- < \lambda < \lambda_+ \tag{8}$$

固有値分布はFig.2に角グラフで示すように,ランダム行列に重なる部分と外れた部分に分かれる. 淡実曲線で(7)式を示し,それにまとわりつく淡角線は正規乱数で計算した結果である. 固有値が3以下の部分はほぼランダム成分と見なせる.3以上の7固有値:$\lambda_1 = 46.2$, $\lambda_2 = 5.25$, $\lambda_3 = 5.04$, $\lambda_4 = 3.90$, $\lambda_5 = 3.51$, $\lambda_6 = 3.41$, $\lambda_7 = 3.11$ が有意成分となる.

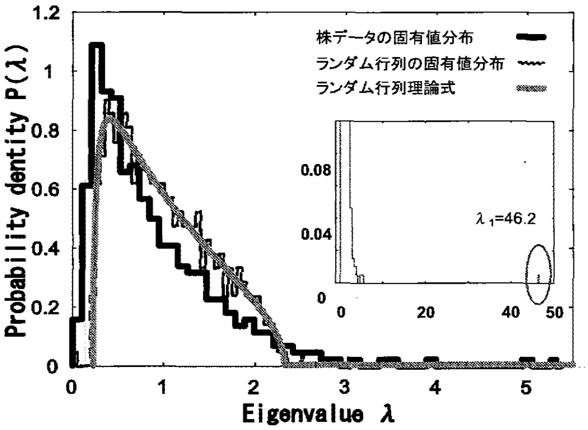

Fig.2 相関行列の固有値(上位 7 個が有意)分布を示す.第 8 固有値以下はランダム成分と見なせる.

次にこの固有値に対する固有ベクトルの成分を調べる.文献[1]では７年分の日次データの解析で, U_1は業種によらず大企業が並んだが,表１に示す今回の解析結果では,1994 年の１時間データの場合もやはり大企業が並ぶので定性的には同じである.但しその企業内訳の詳細はかなり異なる. U_2～U_4に半導体関連企業が多い点も同じである. U_5に石油関連が集中する点は少し違っている.

表 1.固有ベクトル(U_k)成分に見る業種の偏り

U_k	業種成分の偏り
U_1	業種を選ばず大企業数十社 (各 0.08)
U_2	1 位 Barrick Gold(0.3), 2 位以下特徴なし
U_3	10社中7社が半導体に関する企業
U_4	10社中5社が半導体, 他は電気・機械
U_5	10社中9社が石油に関係する企業
U_6	業種に一貫性がない
U_7	紙関連2社,他の業種には一貫性なし
U_8	業種に一貫性がない
U_9	10社中6社が金融関連,2社は通信
U_{10}	自動車2社,通信関連3社

5. 結果の考察と今後の課題

NYSE 株価１時間変動と文献[1]の日次変動の結果との比較を本稿の主目的とした.結果として,第一固有成分に大企業が集まり,第 3, 4 成分に半導体関連が集まるなど,定性的には余り違わない結果となった.今後は 1994 年以外の広範囲のデータでの解析と,１時間毎の代表値をどう選ぶかで違いが出るのか,また異時刻相関の別の扱い方[5]等も検討したい.

6. 参考文献

[1] V. Plerou,et.al.,"Random matrix approach to cross correlation in financial data", Physical Review E 65, 066126, 2002.

[2] M.L.Mehta, "Random Matrices",Academic Press 3rd edition, 2004.

[3] 今野浩,"理財工学 I", 日科技連, 1995.

[4] J.P. Bouchaud, M. Potters, Cambridge,2000: "金融リスクの理論"(森平監訳)朝倉書店,2003.

[5] C.Biely, S. Thurner, Quantitative Finance 8 pp.705-722, 2008.